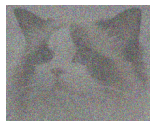
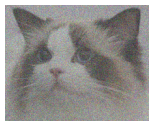


DDPM (2020)  
Theoretical Foundation  
For Image

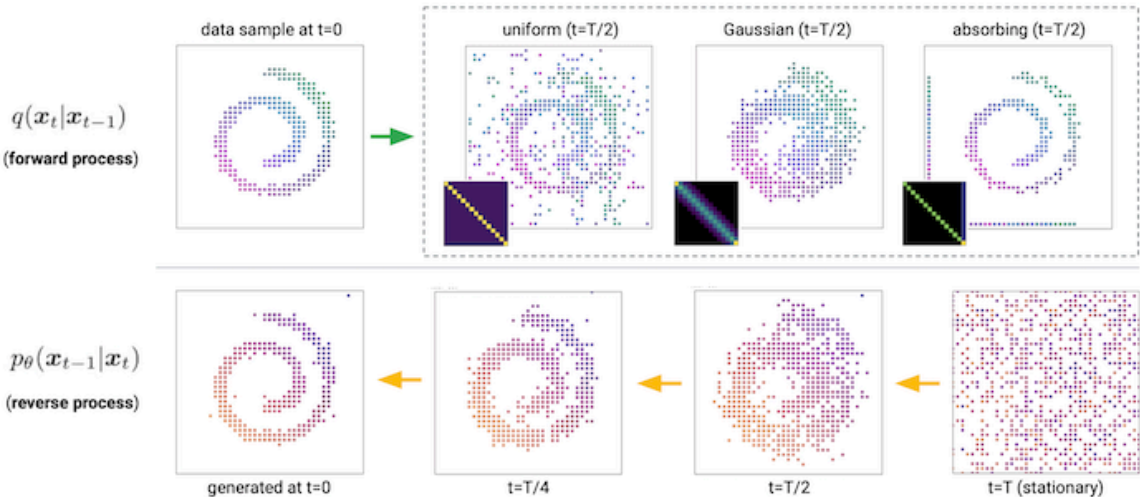
Forward Diffusion Process:

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon$$



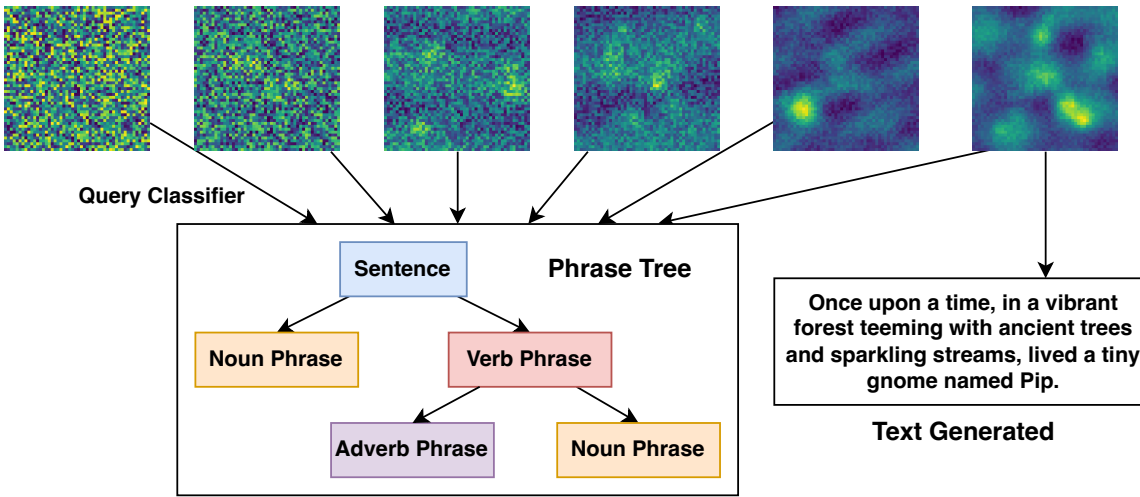
Reverse Diffusion Process: 
$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( x_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_{\theta}(x_t, t) \right) + \sigma_t z$$

D3PM (2021)  
Discrete Breakout






Diffusion-LM (2022)  
Text Adaptation Milestone

Denoise Process on Word Vector



Block Diffusion (2023-2024)  
Hybrid Architecture  
Innovation

Autoregression:		 Generate token for token					
✓	High quality	✓	Arbitrary-length	✓	KV caching	✗	Not Parallelizable
Diffusion:		 Generate globally					
✗	Lower quality	✗	Fixed-length	✗	No KV caching	✓	Parallelizable
Block Diffusion:		 Generate by block					
✓	High quality	✓	Arbitrary-length	✓	KV caching	✓	Parallelizable

LLaDA (2024-2025)  
Large-Scale Pre-training  
Breakthrough

